A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Mobility21 Final Research Report:
Demand learning and supply optimization for last-mile transportation in disadvantaged neighborhoods

Peter Zhang

July 2022

Funding Agency Mobility21, A USDOT National University Transportation Center.

Project Timeline July 1, 2021 - June 30, 2022.

Principal Investigator Peter Zhang, Carnegie Mellon University.

ORCID 0000-0002-0422-834X.

Student Contributors Erin Lin, Cathy Chen, Zhuang Ma, Yidi Miao, Hao Hao, Zehua Wei.

Award Number #368 Demand learning and supply optimization for last-mile transportation in disadvantaged neighborhoods.

FINAL RESEARCH REPORT

Contract # 69A3551747111

Contents

1 Executive Summary

Heritage Community Initiatives (HCI) and their transportation service Heritage Community Transportation (HCT) proposed a transportation problem via the Smart Mobility Challenge in October 2020. The problem was outlined as follows:

> "COVID-19 has reduced the use of public transportation. Despite the relatively low risk of COVID-19 transmission on public transit, implementation of CDC guidelines and publicly available information how passengers can reduce risk of transmission, public transit ridership is down. Rising operating costs with no increase in annual funding, road and bridge construction projects and reduced passenger capacity due to social distancing have made serving the public increasingly difficult. HCT's cost per passenger increased from $12.94 to $24.04 in under six months. We propose a project to analyze HCT service data and public data to ensure that our shuttles are utilizing the best available routes and operating in the most cost-effective, rider-focused manner. Although currently operating as a fixed route line, HCT is also interested analyzing if alternative public transportation methods (deviated fixed route, on-demand pickups, autonomous vehicles, etc.) could better serve our riders."

In this project, we provide extensive analysis of several datasets related to HCT's service region: ridership data from HCT between 2020 and 2022, annual survey data in the past five years, SafeGraph mobility data, Port Authority ridership data, commuting data in Allegheny County, and demographics data. With this descriptive analysis, we provide an in-depth view of the impact of COVID and the future potentials of a transportation service provider in disadvantaged community.

Our findings are as follows.

1. HCT's ridership declined by more than 60% during COVID (Figures 2, 3).
2. Temporary service changes such as shuttle capacity limit due to COVID restrictions, and route changes due to road constructions, are not the causes for the current low level of ridership (Figure 4).
3. HCT's service remained reliable through the pandemic (Figures 5, 6, 7).
4. Demand level is recovering slowly across the HCT service region. Some residents have adopted alternative travel mode (Figure 8).
5. HCT serves residents in neighborhoods with low levels of income and vehicle ownership (Figures 9, 10).
6. The maximum potential economic value of a first-mile and last-mile transportation service is equal to $55.73 per person per day in HCT service region for job related commuting alone (Figures 11, 12, 13). In comparison, based on HCT's own calculation, cost per ride is $12.93 pre-COVID, and $24.04 during COVID.

Given the above evidence and analyses, we believe that HCT provides crucial service for the residents in its service region. In a planned future study (July 2022 to June 2023), we will further examine any additional data and evidence, and provide specific operational and strategic recommendations to modify HCT's service modes.

2 Ridership and Service During COVID-19

Two datasets were provided by Heritage Community Initiatives and its transportation arm, Heritage Community Transportation (HCT).

- Ridership data collected from March 1, 2019 to December 31, 2021.
- Pass-up data collected from June 25, 2020 to June 1, 2021.

The primary data source used to create visualizations throughout this section was the ridership dataset which consists of GPS tracks of vehicles and ridership information in East Pittsburgh, Monroevill, and McKeesport. Basic information such as date, stop time, route, stop location, number of passengers boarding and alighting, number of passengers on the shuttle, and distance to the previous stop were recorded.

In addition to the ridership data, pass-up data is also a specific metric to HCT's service. A pass-up is recorded by the shuttle operator when there are passengers waiting at a stop, but they are unable to board due to capacity limit. Information such as date, stop time, route, stop location, outbound direction, number of passengers passed up, and maximum capacity were recorded.

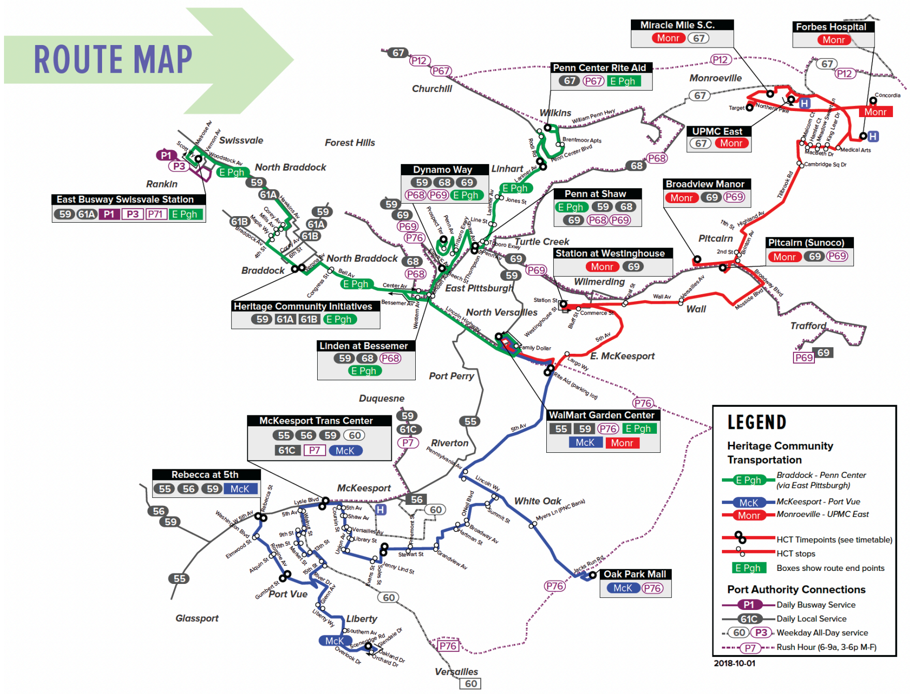

Figure 1: Heritage Community Transportation has three fixed-routes in Allegheny County: McKeesport, Monroeville, and East Pittsburgh. Service map from the Heritage Community Initiatives website (accessed April 2022).

Figure 2 depicts HCT's weekly ridership from 2019 to 2021. The clear drop in March 2020 shows that HCT ridership dropped more than 60% compared with pre-COVID ridership.

In Figure 3, the bubble maps show the spatial distributions of demand, before and after COVID-19 outbreak. The size of the bubble represents the number of riders boarding at a given location. The key observation is that spatial distribution did not shift across locations before and during COVID-19. The magnitude of demand changed across locations almost uniformly.

2.1 Supply Side Analytics

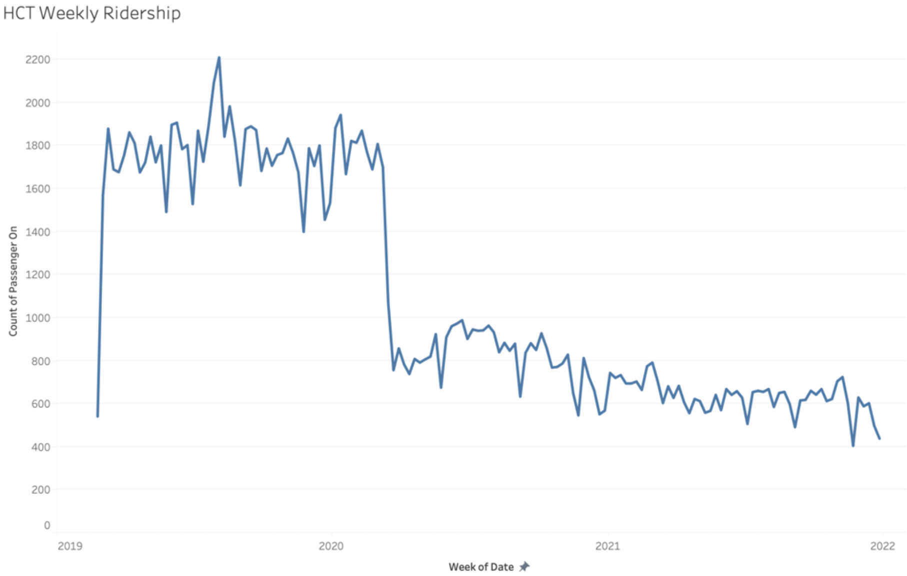

Figure 2: HCT ridership (weekly average) from March 2019 to December 2021.

To understand the sustained low demand level, and its potential causes from the demand side and supply side, we first looked at the supply side of the story: did HCT's service quality and reliability change during COVID? **Overall, HCT provided service in a sustainable and reliable way through COVID.** We illustrate the steps we took to arrive at this conclusion.

Figure 4 demonstrates the change in weekly ridership and major events such as capacity changes or constructions that happened during the time frame. The figure underscores the relationship between the change in ridership and the events. We observe that HCT's shuttle capacity limit in 2020 is unlikely the reason behind the *sustained* low ridership level during COVID-19. The observation can be supported by regression discontinuity analysis around capacity decrease and increase events, as well as weekly passenger count from 2019 to 2021.

In addition, one implication from this analysis is that it is possible that given enough time, ridership is able to bounce back once COVID-19 concerns subside in the future – if we assume residents' work / shopping / appointments patterns resume to pre-COVID style and their transportation choice revert back too. But of course that may not be true, and our study will focus on identifying robust service modes.

Therefore, service changes may have led to short-term ridership fluctuations, but do not explain medium/long-term changes. We further look into more detailed operational statistics to look for the possibility of more granular service change. Our conclusion is that HCT maintained its service level through COVID.

In particular, the length and variability of shuttle's run times are important metrics to evaluate the transit service reliability. Our analyses include the travel time of two scenarios: round trip time and time between any pair of stops. Based on ridership data, we know that shuttles did not have to stop as frequently due to ridership decrease, thus we hypothesized that it could lead to faster travel times and therefore deviate from the shuttle schedule. Analysis shows that this is not the case. HCT has maintained its service reliability through COVID.

Figure 5 and 6 illustrate the distribution of travel times of round trips and a pair of stops, respectively. Surprisingly, the average round trip travel time increases by 2 minute after COVID-19, though the difference may not be statistically significant. On the other hand, travel time between a pair of stops (in this case, Giant Eagle Oat Park and Soles at Stewart on the McKeesport route) remains the same before and after COVID-19.

In addition, we also examined the arrival time of the shuttle at all major stops. The conclusion

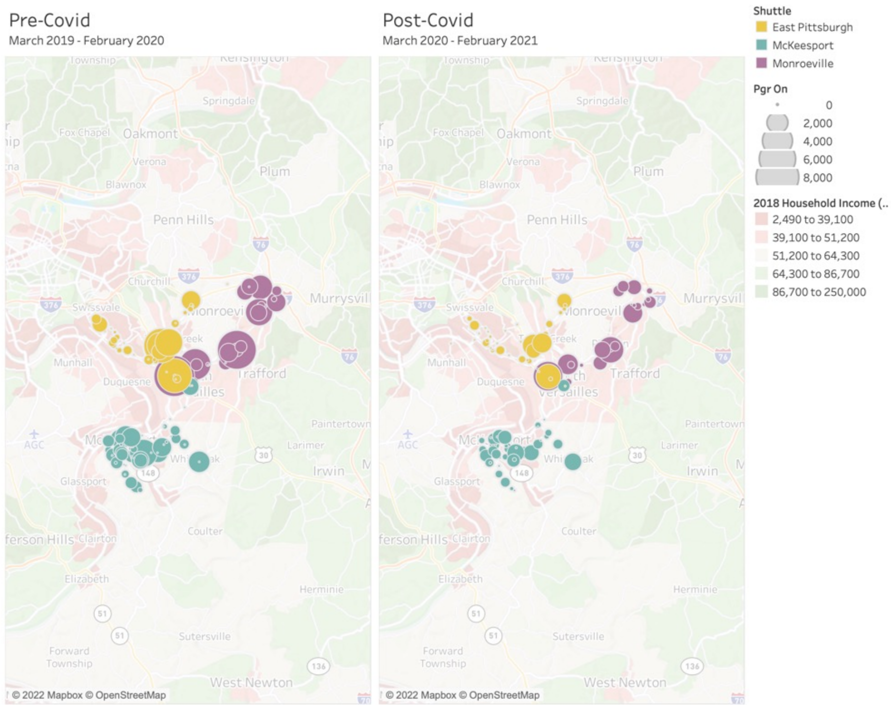

Figure 3: Travel demand distribution across HCT's service area. Spatial demand distribution did not shift, and only decreased in magnitude uniformly.

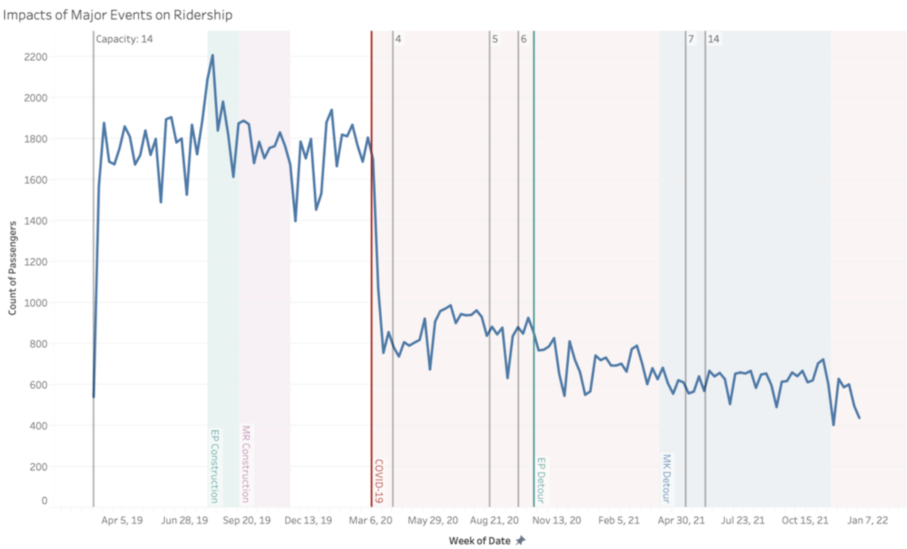

Figure 4: HCT's weekly ridership overlayed with major events. EP refers to the East Pittsburgh route, MK refers to the McKeesport route, MR refers to the Monroeville route. Demand level did not recover after shuttle capacity limit was relaxed, indicating that the travel demand level is still low, and / or the choice of transportation mode has changed.

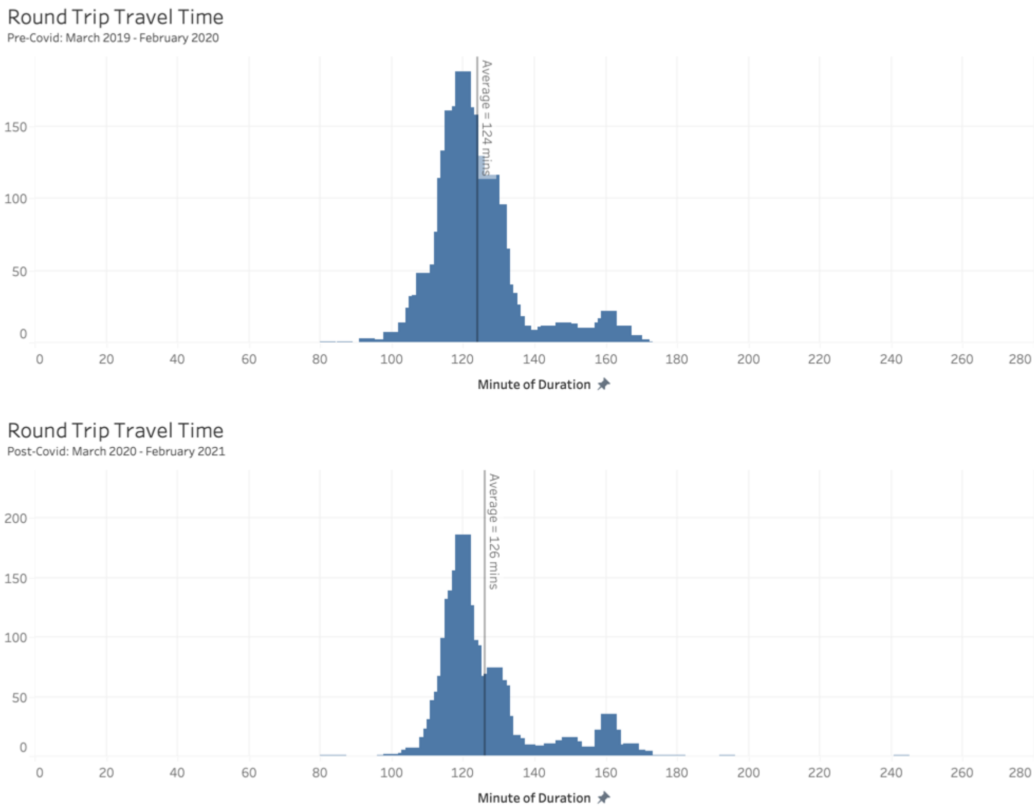

Figure 5: Average round trip travel time remained roughly the same after COVID-19.

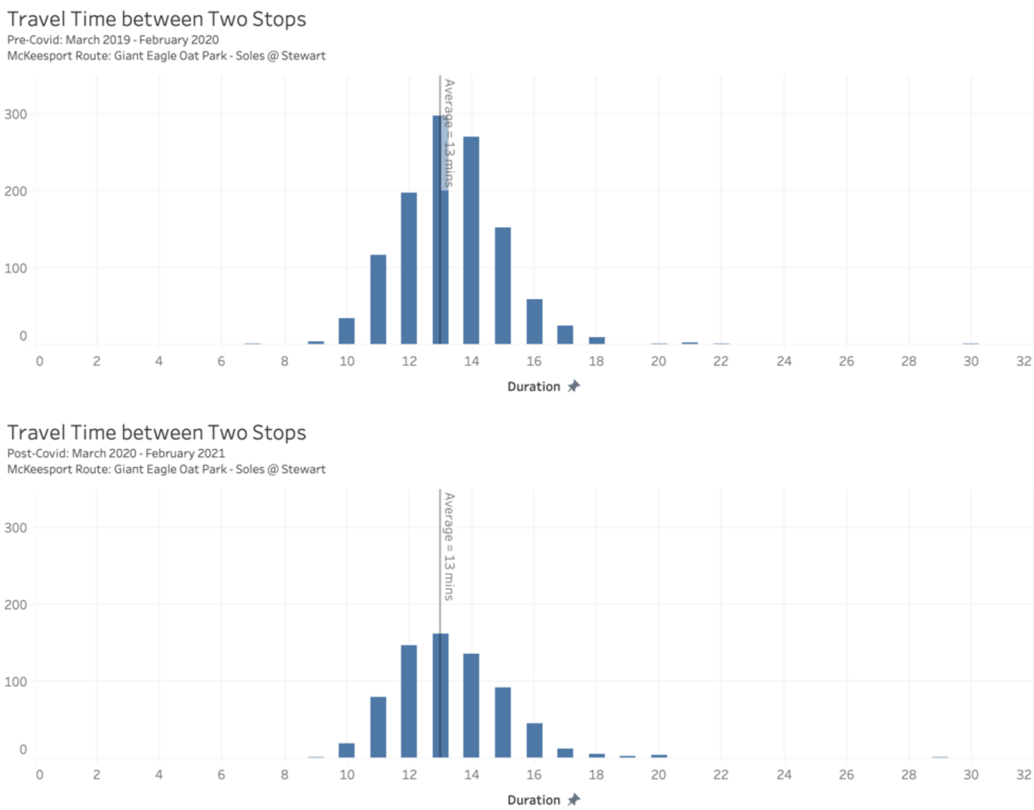

Figure 6: Average travel time between two stops remains the same before and after COVID-19. This plot shows one example: between Giant Eagle Oat Park and Soles at Stewart.

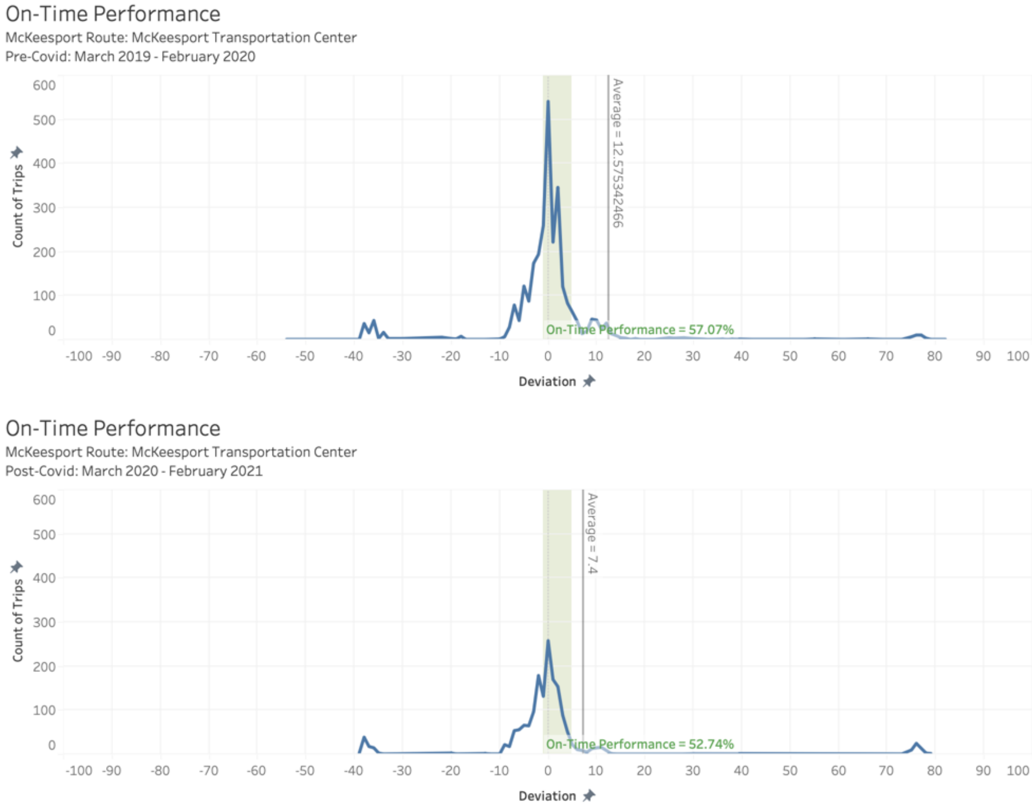

Figure 7: Arrival time distribution at McKeesport Transportation Center. The pattern remained unchanged through COVID.

is the same: arrival times did not change through COVID. Figure 7 illustrates this with the McKeesport Transportation Center stop.

2.2 Demand Side Analytics

Given that HCT's service remained steady and reliable through COVID, yet the demand level is at a sustained low level, we look at the travel demand. There are at least two layers of factors: changes in people's overall demand level, and changes in people's transportation mode choice. **Our conclusion on the demand side is, travel demand has slowly recovered but is significantly below pre-COVID level (by early 2022), and some residents have shifted their transportation choice away from affordable mass transit, and opted to use more expensive but also more flexible and safer individual transit options (jitneys, TNC rides).**

We examined demand level in HCT's service region quantitatively and qualitatively. Quantitatively, we compared the ridership level of HCT with the ambient demand level (measured by SafeGraph's mobility data) and Port Authority's ridership level on the routes that connect with HCT's region. Figure 8 indicates that there is a very slow recovery of travel demand in the region. Therefore, compared with the gradual decline in HCT ridership, we hypothesize that residents have shifted their transportation service choice slightly away from HCT for now.

We looked into survey data from HCT's service region, and observed that indeed some residents have shifted away from HCT, and have chosen to use jitney, Uber/Lyft, and other transportation modes more often. But the trend is not significant.

Thus two questions remain: Given that people are using ridehailing options more often, can

Figure 8: The comparison of HCT ridership, ambient mobility level, and Port Authority ridership. While there is a slow recovery of travel demand, HCT's ridership remained on a slow downward trend. This analysis and additional survey data indicate a shift in riders' transportation choice away from cheap mass transit ($0.25 per rider for HCT) to more expensive and flexible personal transit options such as jitneys and TNC rides.

HCT's service still provide value to its community? If so, should HCT change its service modes to adapt to the demand? We answer the first question with an in-depth top-down analysis of the economic benefits of first-mile and last-mile transportation service. We also propose a research plan for more demand-aware transportation service modes for community transportation provider like HCT.

3 Economic Value of First-Mile and Last-Mile Transportation

In the first subsection, we visualize the income levels and vehicle ownership in the neighborhoods that HCT serves. We show that **HCT's service region include many vulnerable neighborhoods that have low household income and low private vehicle ownership**.

Next, we use data analytics tools (GoogleMap API) to quantify the maximum potential benefit of providing efficient first-mile and last-mile service.

3.1 Income and Mobility Visualization

Figure 9 shows the income level of some neighborhoods around Pittsburgh, focusing on HCT's service region. The dark green dots are HCT shuttle stops. Majority of the neighborhoods served by HCT have an annual household income between \$13,400 and \$50,400.

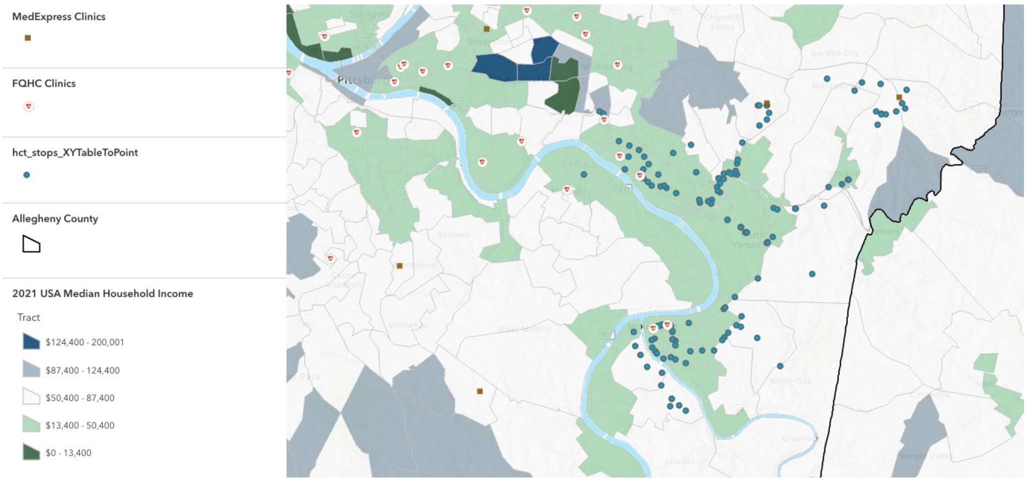

Figure 9: HCT's routes serve residents of low-income neighborhoods. Income and map data from US Census 2020. Map layer created by Esri, accessed April 2022. Overall map created by authors of this report on ArcGIS website.

Figure 10 shows that HCT's service region covers many neighborhoods with high percentage of households that do not have access to vehicles.

3.2 Maximum Potential Benefit for Reducing Commuting Time

In the previous section, we show that residents have low levels of income and access to private vehicles. In this subsection, we further show that they can save time from using first-mile and last-mile transportation service to and from public transit (Port Authority).

In particular, we quantify the maximum potential benefit that a first-mile and last-mile service can provide for these residents in terms of commuting between home and work. We focus on job-related commuting because of data availability. By focusing on this, we do not imply that HCT should focus on job commuting only. To quantify the total benefit that HCT is providing in terms of other activities (e.g., access to health care, grocery shopping, leisure, and education), additional datasets are required and they are not available at the moment.

In the remainder of this section, **we quantify the maximum benefit that a first-mile and last-mile transportation service provider can provide**.

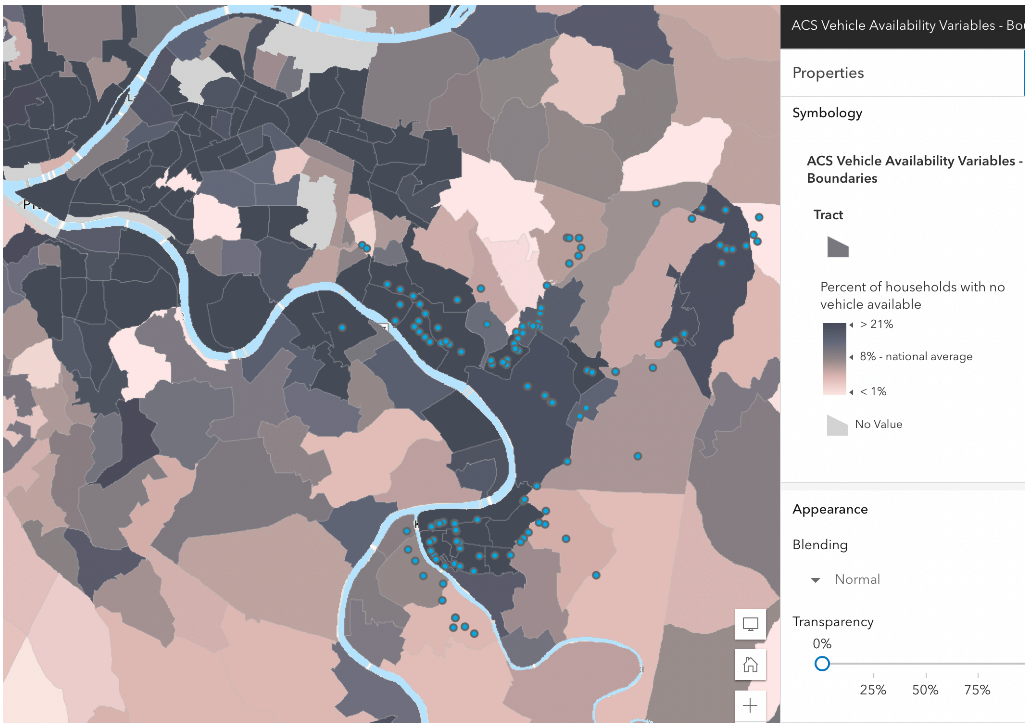

Figure 10: HCT's routes serve residents with fewer vehicles. Vehicle ownership data from American Community Survey, and map layer created by Ersi. Date of API call to the data: March 17, 2022. Overall map created by authors of this report on ArcGIS website.

Assumptions. We make the following assumptions.

- We focus on job related commuting only for this analysis, due to data limitation. (We do not imply HCT would or should only provide job related transportation service).
- We assume that all the first-mile and last-mile services are on-demand, to calculate the maximum benefit of such service.
- We assume that the individuals in the dataset used are representative of HCT's service region. Further demographic and economic distribution information could be found in the bias analysis.

Tools. We used Google Map API to request the possible transit methods, time, costs, and distance of the workers' daily commute. For the consistency of different transit methods, all the workers are set to arrive at their destinations at the same time.

Method. We calculate the economic value of time saving for commuters by the amount of time saved and their hourly salary. Time saved for transit commuters is calculated by the difference between the first-mile and last-mile walking time (requested from Google Map API via walking mode) and the expected travel time if they take on-demand shuttle service (requested from Google Map API via driving mode). In addition, the expected hourly salary is calculated by dividing the average values of the monthly salary recorded in the wages dataset by the average working hours in the United States.

Data. The data we used for this analysis comes from LEHD Origin-Destination Employment Statistics (`https://lehd.ces.census.gov/data/`), or LODES data, a synthetic dataset that describes geographic patterns of jobs by their employment locations and residential locations, as well as the connections between these two locations. Specifically, the data includes three data sources, mainly as follows:

1. The Unemployment Insurance Wage Data, which is reported by employers and maintained by each state to administer its unemployment insurance system, providing information on employees and jobs (relationship between employee and firm).
2. The Quarterly Census of Employment in Wages, which publishes a quarterly count of employment and wages reported by employers.
3. Office of Personnel Management (OPM) - sourced data, which covers more government-related employment information.

By the given definition, a job is counted if a worker is employed with positive earnings during the reference quarter as well as in the quarter prior to the reference quarter. In addition, if a worker is employed at more than one job during the referenced period and the core datasets cover those jobs, then all of those jobs will be captured in the dataset. Besides, these datasets currently exclude several groups of workers: uniformed military, self-employed workers, and informally employed workers.

We look at the data collected in 2019 in Pennsylvania for all job types, from residential locations to employment locations. The total number of records is 5,128,507, and the corresponding features includes the number of jobs in different age groups, income levels, and industries by residential and work locations.

Potential Bias by Using These Datasets. We document the potential bias introduced from using the aforementioned datasets for our analysis. The number of records in the 2019 Pennsylvania origin-destination data files is 5,128,507. Each record represents the number of jobs between a work location and a residential location. By adding up the number of jobs from each record, we could compute the total number of jobs in Pennsylvania to be 5,513,582. If a worker is employed at more than one job during the referenced period and those jobs are covered by the core datasets, then all of those jobs will be captured in the dataset, potentially creating duplicate entries for the same person. Based on the data from the U.S. Bureau of Labor Statistics, the size of the labor force at the end of 2019 is 6,571,438. The data we use in this report represent a large portion of the workforce, but still miss a non-trivial segment that we do not know how to recover.

Economic Value. For the service region that HCT runs through, we estimate that **the maximum benefit that a first-mile and last-mile shuttle service could provide is a saving of 16,002.08 hours for 33,905 workers every day** (Figure 13).

In particular, Figure 11 shows:

- First-mile service could save 32.45 minutes per person per day, for 13,948 workers living in but not working in Monroeville, McKeesport, or East Pittsburgh neighborhoods.
- Last-mile service could save 15.09 minutes per person per day, for 16,790 workers working in but not living in Monroeville, McKeesport, or East Pittsburgh neighborhoods.
- First-mile and last-mile service could save 80.23 minutes for 3,167 workers that both live in and work in Monroeville, McKeesport, or East Pittsburgh neighborhoods.

Based on the pattern, first-mile and last-mile services save the most average commute time for local residents who live and work in the same county, and the most total commute time for those who live in these counties and work outside.

To further quantify the economic value, we borrow data from the Bureau of Labor Statistics (`https://www.bls.gov/emp/tables/output-by-major-industry-sector.htm`, accessed May 2022).

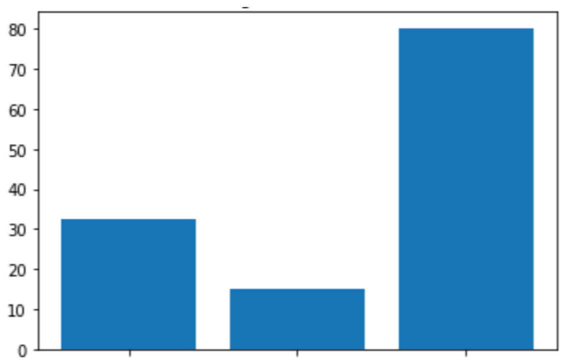

Figure 11: Minutes saved per person per day (y-axis), for those who live in HCT service region but work outside (left column); work in HCT service region but live outside (middle column); and live and work in HCT service region (right column).

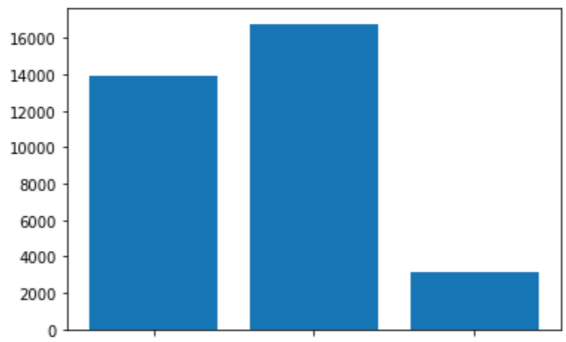

Figure 12: Number of workers (y-axis) separated into three groups: those that live in HCT service region but work outside (left column); work in HCT service region but live outside (middle column); and live and work in HCT service region (right column).

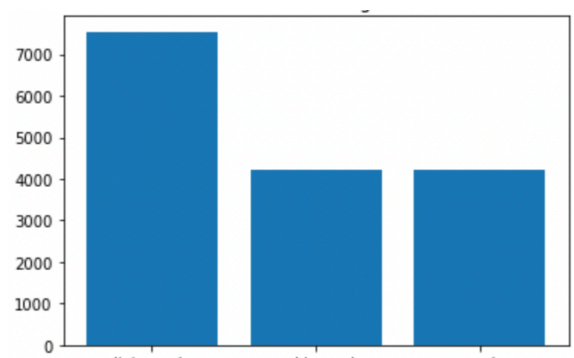

Figure 13: Hours saved per day for all workers (y-axis), for those who live in HCT service region but work outside (left column); work in HCT service region but live outside (middle column); and live and work in HCT service region (right column).

One can show that the economic value created in different sectors to be: Goods Producing industry sectors: \$232.70; trade, Transportation, and Utilities industry sectors: \$210.20; all Other Services industry sectors: \$157.16.

In conclusion, a saving of 16 thousand hours in the HCT service region is the maximum benefit that a first-mile and last-mile shuttle service can support. This translates to \$ 3.78 million dollars per day in economic output, which is equivalent to \$55.73 per person in HCT's service region.

4 Conclusion and Future Work

This current research aims to understand HCT's status and travel demand after the pandemic happened, to provide actionable recommendations. We analyzed different performance metrics that may have impacted the ridership levels. Overall, we find that HCT's service remained stable amid demand decrease. Since the potential economic value of HCT's service is very high, our overall recommendation is to maintain HCT's current service and wait for demand recovery. In addition, HCT may also want to consider a few service modifications to recover its ridership level faster. In a follow-up analysis (July 2022 to June 2023), we will study these potential service modes in more detail with the support of operations research and machine learning methods.

5 Project Output

The following publications and working papers are supported under this grant.

1. Shehadeh, K. S., Wang, H., and Zhang, P. (2021). Fleet sizing and allocation for on-demand last-mile transportation systems. Transportation Research Part C: Emerging Technologies, 132, 103387.
2. Wei, N. and Zhang, P. (2022) Adjustability in Robust Linear Optimization (working paper).
3. Elci, O., Hooker, J., and Zhang, P. (2022) Structural Characteristics and Equitable and Efficient Distributions (working paper).

The following academic conference presentations are supported by this grant.

1. "Adjustability in Robust Linear Optimization", INFORMS Optimization Society Conference, March 2022, Greenville SC.
2. "Demand learning and supply optimization for last mile transportation in low-income neighborhood", POMS Conference 2022, April 2022, virtual.
3. "Adjustability in Robust Linear Optimization", International Conference on Continuous Optimization, July 2022, Bethlehem PA.

The following dataset is curated to support the analysis of commuting times and transportation equity:

1. U.S. Household Commuting Dataset and Transportation Fairness. Dataset: https://github.com/peteryz/employment-od. Data sheet: https://ppms.cit.cmu.edu/media/project_files/Mobility_Fairness_Dataset.pdf.